

# Deformation and Illumination Invariant Feature Point Descriptor

Francesc Moreno-Noguer  
Institut de Robòtica i Informàtica Industrial (CSIC-UPC)  
08028, Barcelona, Spain

## Abstract

*Recent advances in 3D shape recognition have shown that kernels based on diffusion geometry can be effectively used to describe local features of deforming surfaces. In this paper, we introduce a new framework that allows using these kernels on 2D local patches, yielding a novel feature point descriptor that is both invariant to non-rigid image deformations and illumination changes.*

*In order to build the descriptor, 2D image patches are embedded as 3D surfaces, by multiplying the intensity level by an arbitrarily large and constant weight that favors anisotropic diffusion and retains the gradient magnitude information. Patches are then described in terms of a heat kernel signature, which is made invariant to intensity changes, rotation and scaling. The resulting feature point descriptor is proven to be significantly more discriminative than state of the art ones, even those which are specifically designed for describing non-rigid image deformations.*

## 1. Introduction

Building invariant region descriptors is a major topic of research in computer vision with a wide range of applications in areas of object recognition and 3D reconstruction. Over the last decade, great success has been achieved in designing descriptors invariant to certain types of geometric and photometric transformations. For instance, the SIFT [16] descriptor and many of its variants [1, 10, 17, 24] have been proven invariant to affine deformations of both spatial and intensity domains. In addition, affine deformations can effectively approximate, at least on a local scale, other image transformations including perspective and viewpoint changes. However, this approximation is no longer valid for general deformations as those occurring when viewing an object that deforms non-rigidly [26].

In order to match points of interest under non-rigid image deformations, recent approaches propose optimizing complex objective functions that ensure global consistence

among the spatial layout of all matches [8, 9, 11, 21, 25]. Yet, none of these approaches explicitly builds a descriptor that goes beyond affine transformations. An interesting exception is [13], that proposes embedding the image as a 3D surface and using a Geodesic Intensity Histogram (GIH) as a feature point descriptor. However, while this approach is robust to non-rigid deformations, its performance drops under light changes. This is because the GIH considers deformations as one-to-one mappings where image pixels only change their position but not their intensities.

To overcome the inherent limitation of a metric that uses GIH distances, we propose a novel descriptor based on the Heat Kernel Signature (HKS) recently introduced for non-rigid 3D shape recognition [20, 23], and which besides invariance to deformation, it has been demonstrated to be invariant to global isotropic scalings [5]. This is particularly interesting in our context of images embedded on 3D surfaces, because illumination changes produce variations on the intensity dimension that can be seen as local anisotropic scalings, for which [5] still shows a good resilience.

Our main contribution is thus in using the tools of diffusion geometry to build a descriptor for 2D image patches invariant to non-rigid deformations and photometric changes. More specifically, we consider an intensity image  $I$  as a surface in the  $(x, y, \beta I)$  space, where  $(x, y)$  are the spatial coordinates and  $\beta$  is a parameter which is set to an arbitrarily large value to ensure the descriptor preserves edge information. Drawing motivation from the HKS [5, 23], we then describe each feature point in terms of the heat it dissipates onto its neighborhood over time. To increase robustness against 2D and intensity noise, we consider the heat diffusion process not only on the point of interest, but also on the pixels of its neighborhood, accordingly weighed by a Gaussian kernel. As shown in Fig. 1, the resulting descriptor (which we call DaLI, for Deformation and Light Invariant) outperforms state-of-the-art descriptors in matching points of interest between images that have undergone non-rigid deformations and photometric changes.

## 2. Related Work

The SIFT descriptor [16] has become the main reference among keypoint descriptors, showing great success in cap-

---

This work has been partially funded by the Spanish Ministry of Science and Innovation under projects DPI2008-06022 and Consolider Ingenio 2010 CSD2007-00018; and by the EU project GARNICS FP7-247947.

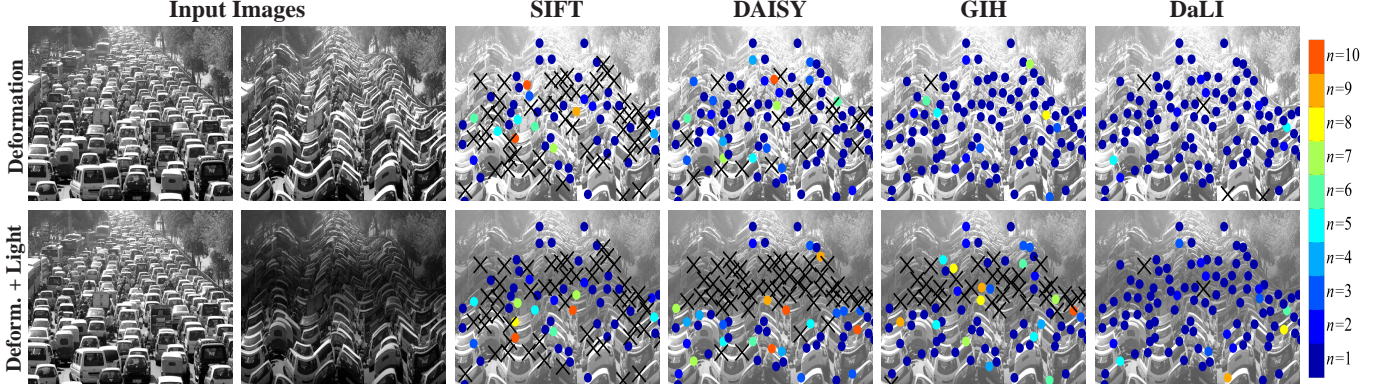


Figure 1. Comparing DaLI against other descriptors. **Top: Matching under Non-Rigid Deformations.** We compare DaLI, against SIFT [16], DAISY [24] and GIH [13]. Colored circles indicate the match has been correctly found among the first  $n$  top candidates, where  $n \leq 10$  is parameterized by the legend on the right. A feature is considered as mismatched when  $n > 10$  and we indicate this with a cross. Note that GIH, which is specifically designed for handling non-rigid deformations, yields very good results, similar to our approach. **Bottom: Matching under Non-Rigid Deformations and Light Changes.** Under this situation, only our approach remains robust.

turing local affine deformations including scaling, rotation and certain lighting changes. Since it is relatively slow to compute, most of the subsequent works have focused in developing faster descriptors [1, 6, 10, 17, 24]. However, little effort has been dedicated in building descriptors robust to more general deformations.

The limitations of the affine-invariant descriptors when solving correspondences under non-rigid deformations are compensated by enforcing global consistency, both spatial and photometric, among all the features [2, 3, 8, 9, 11, 21, 25]. Thus, all these methods do not handle the non-rigid nature of the problem, and rely on solving complex optimization functions for establishing matches.

An alternative approach, is to directly build a deformation invariant descriptor. With that purpose, recent approaches in two-dimensional shape analysis have proposed using different types of intrinsic geometry. For example, [4, 14] define metrics based on the inner-distance, and [15] proposes using geodesic distances. However, all these methods require the shapes to be segmented out and represented by binary images, which is difficult to hold in practice. In [13], it was shown that geodesic distances, in combination with an appropriate 3D embedding of the image, were adequate to achieve deformation invariance in intensity images. Yet, this method assumes that pixels only change its position and not its intensity and, as shown in Fig. 1, becomes unreliable under illumination changes.

Motivated by [13], we represent the images as 2D surfaces embedded in the 3D space. In fact, this is quite an old idea, although it has been mostly employed for low level vision tasks such as image denoising [22, 29] and segmentation [28]. The fundamental difference between our approach and [13] is that we then describe each feature point on the embedded surface considering the heat diffusion over time [5, 23] instead of using a Geodesic Intensity

Histogram. As we will show in the results section this yields substantially improved robustness, especially to lighting changes. Again, the heat diffusion theory has been used in several early approaches for 3D shape analysis [12, 19, 20], but to the best of our knowledge, it has never been used for describing image patches in intensity images.

### 3. Deformation and Light Invariant Descriptor

Our approach is inspired by current methods [5, 23] that suggest using the diffusion geometry for 3D shape recognition. In this section we show how this theory may be adapted for describing 2D local patches of images that undergo non-rigid deformations and photometric changes.

#### 3.1. Invariance to Non-Rigid Deformations

Let us assume we want to describe a 2D image patch  $P$ , of size  $S \times S$  and centered on a point of interest  $\mathbf{p}$ . In order to apply the diffusion geometry theory on intensity patches we regard them as 2D surfaces embedded in 3D space (Fig. 2 Top-left). More formally, let  $f : P \rightarrow M$  be the mapping of the patch  $P$  to a 3D Riemannian manifold  $M$ . We explicitly define this mapping by:

$$f : \mathbf{x} \rightarrow (x, y, \beta P(\mathbf{x})) \quad \forall \mathbf{x} \in P, \quad (1)$$

where  $P(\mathbf{x})$  is the pixel intensity at  $\mathbf{x} = (x, y)^\top$ , and  $\beta$  is a parameter that, as we will discuss later, controls the amount of gradient magnitude preserved in the descriptor.

Several recent methods [12, 19, 20, 23] have used the heat diffusion geometry for capturing the local properties of 3D surfaces and performing shape recognition. Along these lines, we describe each patch  $P$  based on the heat diffusion equation over the manifold  $M$ :

$$\left( \Delta_M + \frac{\partial}{\partial t} \right) u(\mathbf{x}, t) = 0, \quad (2)$$

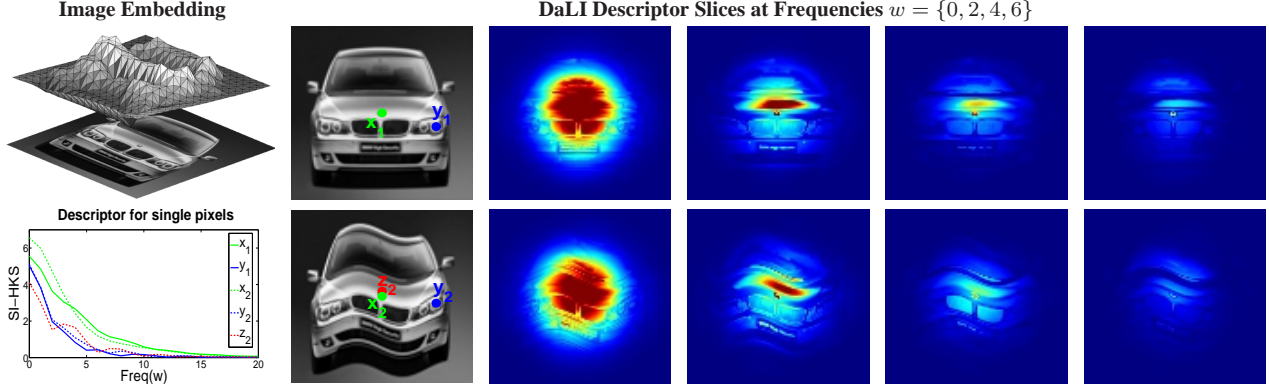


Figure 2. DaLI descriptor. **Top Left:** Our central idea is to embed image patches on 3D surfaces and describe them based on heat diffusion processes. We represent the heat diffusion as a stack of images in the frequency domain. **Right:** Various slices of our descriptor for two different patches. **Bottom Left:** Value of the descriptor for the pixels marked by color circles in the images on the right. Note that corresponding pixels have very similar signatures. However, the signature may change from one pixel to its immediate neighbor. For instance,  $\mathbf{z}_2$  is at one pixel distance from  $\mathbf{x}_2$ , but their signature is rather different. As a consequence, using the signature of a single point as a descriptor is prone to be very sensitive to 2D noise in the feature detection process. We address this issue by simultaneously considering the signature of all the pixels within the patch, weighted by a Gaussian function of the distance to the center of the patch.

where  $\Delta_M$  is the *Laplace-Beltrami operator*, a generalization of the Laplacian to non-Euclidean spaces. The solution  $k(\mathbf{x}, \mathbf{y}, t)$  of the heat equation is called the *heat kernel*, and represents the amount of heat that is diffused between points  $\mathbf{x}$  and  $\mathbf{y}$  at time  $t$ , considering a unit heat source at  $\mathbf{x}$  at time  $t = 0$ . For a compact manifold  $M$ , the heat kernel can be expressed by following spectral expansion [7, 19]:

$$k(\mathbf{x}, \mathbf{y}, t) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(\mathbf{x}) \phi_i(\mathbf{y}), \quad (3)$$

where  $\{\lambda_i\}$  and  $\{\phi_i\}$  are the eigenvalues and eigenvectors of  $\Delta_M$ . Based on this expansion, [23] proposes describing a point  $\mathbf{p}$  on  $M$  using the Heat Kernel Signature

$$\text{HKS}(\mathbf{p}, t) = k(\mathbf{p}, \mathbf{p}, t) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i^2(\mathbf{p}), \quad (4)$$

which is shown to be isometrically-invariant, and adequate for capturing both the local properties of the shape around  $\mathbf{p}$  (when  $t \rightarrow 0$ ) and the global structure of  $M$  (when  $t \rightarrow \infty$ ).

However, while on smooth surfaces the HKS of neighboring points are expected to be very similar, when dealing with the irregular shapes that may result from embedding image patches, the heat kernel turns to be highly unstable along the spatial domain (Fig. 2 Bottom-left). This makes the HKS particularly sensitive to noise in the 2D coordinates of the keypoints. To handle this situation, we build the descriptor of a point  $\mathbf{p}$  by assembling the HKS of all points  $\mathbf{x}$  within the patch  $P$ , properly weighted by a Gaussian function of the distance to the center of the patch. Thus, we define the following Deformation Invariant (DI) descriptor:

$$\text{DI}(\mathbf{p}, t) = [\text{HKS}(\mathbf{x}, t) \cdot G(\mathbf{x}; \mathbf{p}, \sigma)]_{\forall \mathbf{x} \in P}, \quad (5)$$

where  $G(\mathbf{x}; \mathbf{p}, \sigma)$  is a 2D Gaussian function with mean  $\mathbf{p}$  and standard deviation  $\sigma$ , evaluated at  $\mathbf{x}$ .

The price we pay for achieving increased robustness to 2D noise is an additional complexity of the descriptor. That is, if  $\text{HKS}(\mathbf{p}, t)$  is a function defined on the temporal domain  $\mathbb{R}^+$ , the new descriptor  $\text{DI}(\mathbf{p}, t)$  is defined on  $S \times S \times \mathbb{R}^+$ , the product of the spatial and temporal domains. However, note that for our purposes this is still feasible, because we do not need to compute a descriptor for every pixel of the image, but just for a few hundreds of points of interest. Furthermore, as we will next discuss, the descriptor may be highly compacted if we represent it in frequency domain instead of time domain.

### 3.2. Invariance to Illumination Changes

An intrinsic limitation of the descriptor introduced in Eq. 5 is that it is not illumination invariant. This is because light changes, scale the manifold  $M$  along the intensity axis, and HKS is sensitive to scaling. Indeed, it can be shown that an isotropic scaling of the manifold  $M$  by a factor  $\alpha$ , scales the eigenvectors and eigenvalues of Eq. 3 by factors  $1/\alpha$  and  $1/\alpha^2$ , respectively [19]. The HKS of a point  $\alpha\mathbf{p} \in \alpha M$  can then be written as

$$\text{HKS}(\alpha\mathbf{p}, t) = \sum_{i=0}^{\infty} e^{-\frac{\lambda_i}{\alpha^2} t} \frac{\phi_i^2(\mathbf{p})}{\alpha^2} = \frac{1}{\alpha^2} \text{HKS}(\mathbf{p}, \frac{t}{\alpha^2}), \quad (6)$$

which is an amplitude and time scaled version of the original HKS.

Nonetheless, under isotropic scalings, several alternatives have been proposed to remove the dependence of HKS to the scale parameter  $\alpha$ . For instance, [19] suggests normalizing the eigenvalues in Eq. 3. [5] proposes applying three consecutive transformations on the HKS. First, the



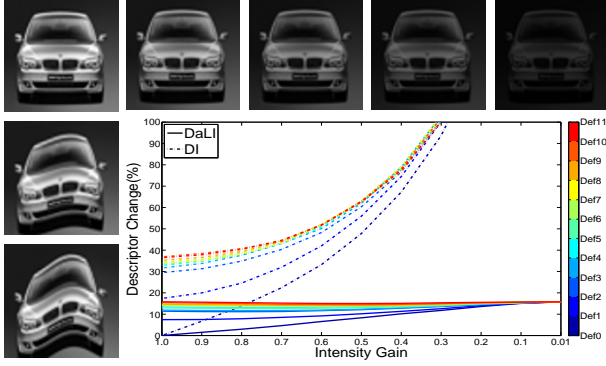


Figure 3. Invariance of the DaLI and DI descriptors to non-rigid deformations and illumination changes. **Top and Left Images:** Different degrees of deformation and light changes applied on the top left reference patch  $P_0$ . Deformations are applied according to a function  $\text{Def}(\cdot) \in \{\text{Def0}, \dots, \text{Def11}\}$ , where  $\text{Def11}$  corresponds to the maximal deformation. Light changes are produced by scaling the intensity of  $P_0$  by a gain  $g \in [0, 1]$ . **Bottom Graph:** Given a deformation  $\text{Def}(\cdot)$  and a gain factor  $g$ , we compute the percentage of change of the DI descriptor by  $\|DI(P_0) - DI(\text{Def}(gP_0))\| / \|DI(P_0)\|$ . The percentage of change for DaLI is computed in a similar way. Observe that the DaLI is much less sensitive than DI, particularly to lighting changes.

time-dimension is logarithmically sampled, which turns the time scaling into a time-shift, that is, the right-hand side of Eq. 6 begets  $\alpha^{-2} \text{HKS}(\mathbf{p}, -2 \log \alpha + \log t)$ . Second, the amplitude scaling factor is removed by taking the logarithm and a derivative w.r.t.  $\log t$ . The Heat kernel then becomes  $\frac{\partial}{\partial \log t} \text{HKS}(\mathbf{p}, -2 \log \alpha + \log t)$ . The time-shift term  $-2 \log \alpha$  is finally removed using the magnitude of the Fourier transform, which yields  $\text{SI-HKS}(\mathbf{p}, w)$ , a scale invariant version of the original HKS in the frequency domain. In addition, since most of the signal information is concentrated in the low-frequency components, the complexity of the descriptor may be highly reduced compared to that of the  $\text{HKS}(\mathbf{p}, t)$ .

Another advantage of the SI-HKS signature is that although it is specifically designed to remove the dependence of the HKS to isotropic scalings, it is quite resilient to anisotropic transformations, such as those produced by intensity changes, that only affect the intensity dimension of the manifold  $M$ . Thus, we will use this signature to define our Deformation and Light Invariant (DaLI) descriptor<sup>1</sup>:

$$\text{DaLI}(\mathbf{p}, w) = [\text{SI-HKS}(\mathbf{x}, w) \cdot G(\mathbf{x}; \mathbf{p}, \sigma)]_{\forall \mathbf{x} \in P} \quad (7)$$

Fig. 2 shows several slices of the DaLI descriptor at different frequencies for a patch and a deformed version of it. As said above, observe that most of the signal is concentrated on the low frequency components. In Fig. 3 we compare the sensitivity of the DI and DaLI descriptors to defor-

<sup>1</sup>In the rest of the paper we will indistinctly use  $\text{DaLI}(\mathbf{p}, w) \equiv \text{DaLI}(\mathbf{p}) \equiv \text{DaLI}(P)$ .

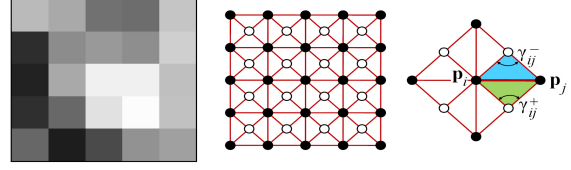


Figure 4. **Left:** Image patch. **Center:** Representation of the patch as a triangular mesh. For clarity of presentation we only depict the  $(x, y)$  dimension of the mesh. Note that besides the vertices placed on the center of the pixels (black circles) we have introduced additional intra-pixel vertices (white circles), that provide finer heat diffusion results and higher tolerance to in-plane rotations. **Right:** Definition of the angles used to compute the discrete Laplace-Beltrami operator.

mation and light changes. Note that DaLI, in contrast DI, remains almost invariant to light changes, and it also shows a better performance under deformations. In the results section, we will show that this invariance is also accompanied by a high discriminability, yielding significantly better results in keypoint matching than existing approaches.

### 3.3. Handling in-Plane Rotation and Scaling

Although DaLI tolerates certain amounts of in-plane rotation and scaling, it is not designed for these purposes. This is because with the aim of increasing robustness to 2D noise, we built the descriptor using all the pixels within the patch, and their spatial relations have been retained. Thus, if the patch is rotated or scaled, so will be the descriptor.

In order to handle this situation, during the matching process we will consider several rotated and scaled copies of the descriptors. Therefore, given  $\text{DaLI}(\mathbf{p}_1)$  and  $\text{DaLI}(\mathbf{p}_2)$  we will compare them based on the following metric

$$d(\mathbf{p}_1, \mathbf{p}_2) = \arg \min_{\{\theta_i, s_j\}} \|T_{\theta_i, s_j}(\text{DaLI}(\mathbf{p}_1)) - \text{DaLI}(\mathbf{p}_2)\| \quad (8)$$

where  $\|\cdot\|$  denotes the  $L_2$  norm and  $T_{\theta_i, s_j}(\text{DaLI}(\mathbf{p}))$  rotates and scales  $\text{DaLI}(\mathbf{p})$  by an angle  $\theta_i$  and a scale parameter  $s_j$ , respectively. These parameters are chosen among a discrete set of values  $\theta$  and  $s$ .

### 3.4. Implementation Details

We next describe a number of important details to be considered for the implementation of the DaLI descriptor.

#### Geometry of the embedding

For the numerical computation of the heat diffusion, it is necessary to discretize the problem. Thus, the manifold  $M$  on which the image patch is embedded is represented by a triangulated mesh. Since it is known that the way the heat diffusion behaves depends on the geometry of this mesh [27], we examined different triangulations: structured with either 6 or 8 neighbors per vertex, and unstructured. Fig. 4 shows the representation we use for all the experiments. Although it requires introducing additional vertices

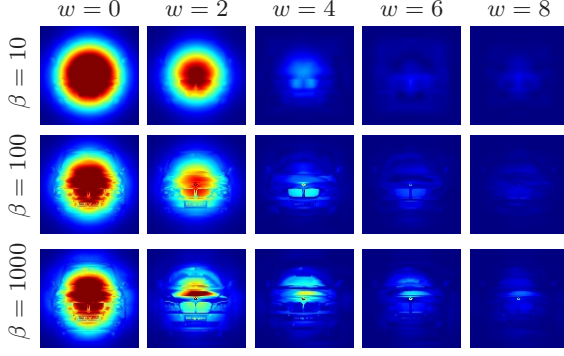


Figure 5. Preserving edge information. Larger values of the parameter  $\beta$  in Eq. 1 allow the descriptor to retain edge information. Each row depicts the DaLI descriptor at frequencies  $w = 0, 2, 4, 6, 8$  for a different value of  $\beta$ .

between the pixels, its symmetry with respect to the  $x$  and  $y$  directions provides robustness to small amounts of rotation, and more uniform diffusions than other configurations.

Another important variable of our design is the magnitude of the parameter  $\beta$  in Eq. 1, that controls the importance of the intensity coordinate with respect to the  $(x, y)$  coordinates. In particular, as shown in Fig. 5, large values of  $\beta$  allow our descriptor to preserve edge information. This is a remarkable feature of the DaLI descriptor, because besides being deformation and illumination invariant, edge information is useful to discriminate among different patches.

#### Discretization of the Laplace-Beltrami operator

In order to approximate the Laplace-Beltrami eigenfunctions on the triangular mesh we use the cotangent scheme described in [18]. We next detail the main steps.

Let  $\{\mathbf{p}_1, \dots, \mathbf{p}_{n_v}\}$  be the vertices of a triangular mesh, associated to an image patch embedded on a 3D manifold. We approximate the discrete Laplacian by a  $n_v \times n_v$  matrix  $\mathbf{L} = \mathbf{A}^{-1}\mathbf{M}$  where  $\mathbf{A}$  is a diagonal matrix in which  $\mathbf{A}_{ii}$  is proportional to the area of all triangles sharing the vertex  $\mathbf{p}_i$ .  $\mathbf{M}$  is a sparse matrix computed by:

$$\mathbf{M}_{ij} = \begin{cases} \sum_k m_{ij} & \text{if } i = j \\ -m_{ij} & \text{if } \mathbf{p}_i \text{ and } \mathbf{p}_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $m_{ij} = \cot \gamma_{ij}^+ + \cot \gamma_{ij}^-$ , and  $\gamma_{ij}^+$  and  $\gamma_{ij}^-$  are the two opposite angles depicted in Fig.4-Right. The eigenvectors and eigenvalues of the discrete Laplace-Beltrami operator may then be computed from the solution of the generalized eigenproblem  $\mathbf{M}\Phi = \mathbf{A}\Lambda\Phi$ , where  $\Lambda$  is a diagonal matrix, in which the eigenvalues  $\{\lambda_i\}$  and the columns of  $\Phi$  correspond to the eigenvectors  $\{\phi_i\}$  in Eq. 3.

## 4. Experimental Results

In this section we compare DaLI against other descriptors for matching points of interest in images that have undergone non-rigid deformations and illumination changes.

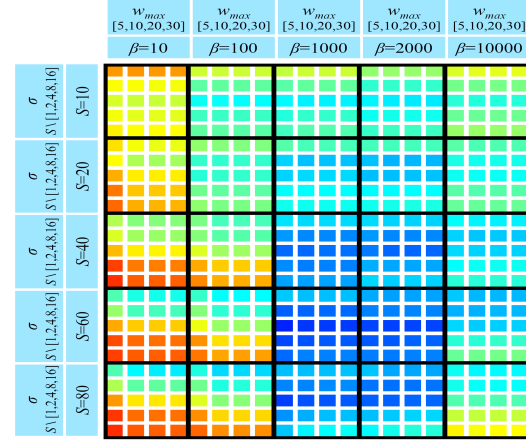


Figure 6. DaLI performance for different values of the parameters  $S$ ,  $\beta$ ,  $\sigma$  and  $w_{max}$ . We compute and average the matching rate for the two pairs of images shown in Fig. 1 using  $S \in \{10, 20, 40, 60, 80\}$  pixels,  $\beta \in \{10, 10^2, 10^3, 2 \times 10^3, 10^4\}$ ,  $\sigma \in \{\frac{S}{16}, \frac{S}{8}, \frac{S}{4}, \frac{S}{2}, S\}$  and  $w_{max} \in \{5, 10, 20, 30\}$ . The graph depicts the results of this 4D parameter exploration, where the color of each square represents the percentage of mismatched points for a specific combination of the parameters.

### 4.1. Choosing Descriptor's Parameters

Before doing the comparison we will take a closer look at the parameters that control the performance of our descriptor, namely the size  $S$  of the patch, the magnitude  $\beta$  of the embedding, the degree  $\sigma$  of smoothing within the patch, and the dimensionality  $w_{max}$  of the descriptor in the frequency domain. In order to find their optimal values, we used the images shown in Fig. 1, and computed mismatch rates for a wide range of values for each of the parameters.

The results of this parameter sweeping experiment are shown in Fig. 6. It can be seen that the most influential parameters are the size of the patch and the magnitude of the embedding. In particular, the best matching results are obtained for  $40 \geq S \geq 60$  and  $1000 \geq \beta \geq 2000$ . Note that these levels of  $\beta$ , ensure that the image gradient is retained by the descriptor. In fact, this was already expected, as the gradient is an important clue for recognition. On the other hand, the influence of the smoothing parameter  $\sigma$  and especially of the number of components necessary is rather limited. Yet, Fig. 6 still shows some pattern for these parameters, such that the best results are obtained for  $\frac{S}{8} \geq \sigma \geq \frac{S}{4}$  and  $10 \geq w_{max} \geq 30$ . Based on this analysis, in all the experiments we report in the next section we set  $S = 60$ ,  $\beta = 2000$ ,  $\sigma = \frac{S}{4}$  and  $w_{max} = 20$ . In addition, by setting the rotation and scaling parameters of Eq. 8 to  $\phi = \{-20, 10, 0, 10, 20\}$  and  $\mathbf{s} = \{0.8, 1, 1.2\}$ , we allow descriptor rotations of 40 degrees and scalings up to 20%.

### 4.2. Comparison with Other Approaches

We compare the performance of our descriptor to that of the SIFT [16], DAISY [24] and GIH [13]. The first

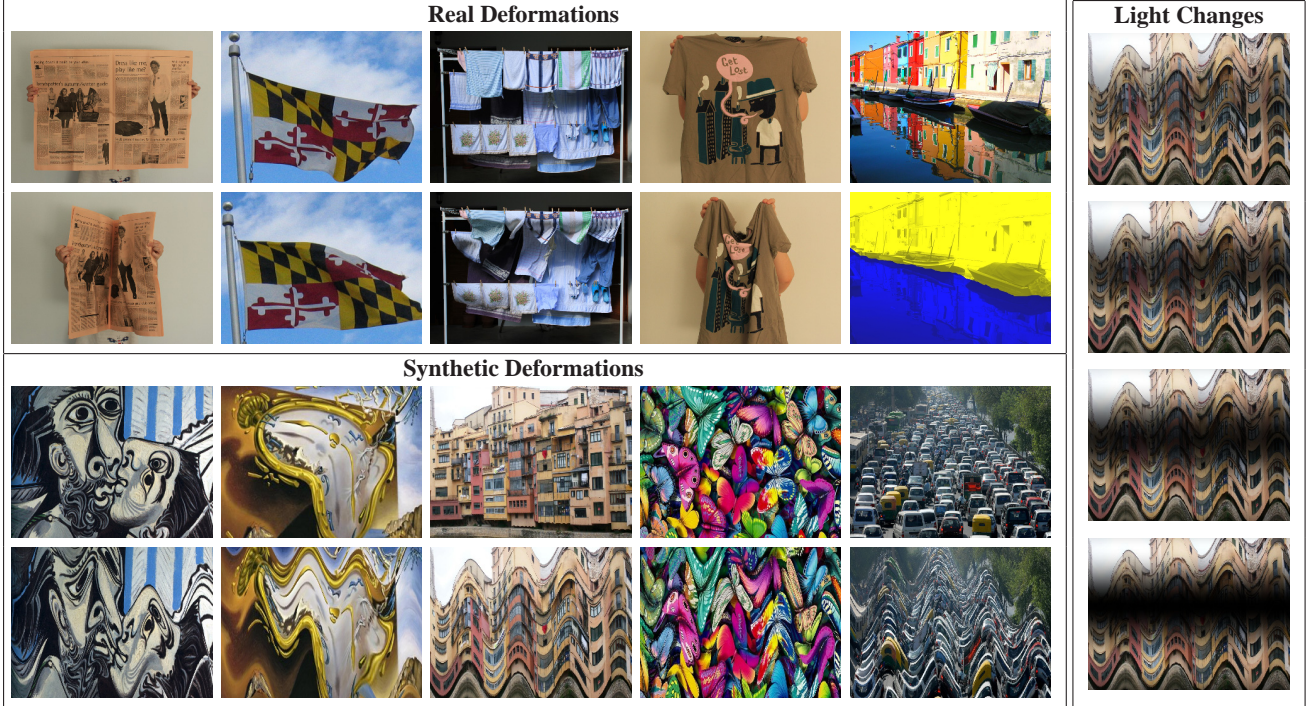


Figure 7. **Left:** Sample images from the test set used for comparison. The dataset includes both synthetic and real deformations, and considers different situations and levels of complexity, from the sharp deformations of a piece of cloth under wind to the smooth distortions produced when an object is reflected on the water. In this case, we manually segmented the area of the image belonging to the object from its reflection and performed matching between these two regions (Top-right). For the synthetic deformations we made sure to produce highly non-linear transformations, in order to assess the robustness of the descriptors to non-affine deformations. **Right:** For evaluating the method to light changes, the intensity of the images was non-linearly transformed as a sinusoidal function of increasing amplitude. In the following we denote them as Light Condition  $\#i$  where  $i \in \{0, 1, 2, 3, 4\}$ . The case  $i = 0$  implies no light change.

two, are representative examples of local region descriptors designed for robustness to affine deformations and certain lighting changes. The GIH, instead, is specifically designed to handle non-rigid image deformations, but as pointed previously, it assumes these deformations are the result of changing the position of the pixels within the image, but not their intensity. This makes GIH sensitive to light changes. In order to address this issue, [13] proposes introducing a scale parameter and compute the GIH descriptor for different magnitudes of the intensity domain. When small light changes are expected this parameter is assumed to change within the interval  $[0.95, 1.05]$ . Since we want to deal with relatively large illumination changes, we also consider the case that this parameter changes within the interval  $[0.25, 1.75]$ . The GIHs computed in each case will be denoted by  $\text{GIH}[0.95, 1.05]$  and  $\text{GIH}[0.25, 1.75]$ .

For evaluating the descriptors, we acquired a dataset of non-rigidly deformed pairs of images, of  $640 \times 480$  pixels, containing 10 real and 11 synthetic deformations. In addition, for each pair of images, the intensity of one of them was progressively modified (in 5 steps) in order to simulate photometric variations. As a result, the whole database is made of 105 image pairs under different levels of deforma-

tion and photometric changes. Fig. 7 shows a few samples.

Given an image pair, we then compute the recognition rate of each descriptor as follows: (1) Using the SIFT detector, we pick  $100 \geq N \geq 200$  points of interest in the first image, and estimate their ground truth position in the second image. (2) We then compute the descriptor for all keypoints in the two images. (3) Based on the descriptor-specific metrics, we estimate the  $n$  top matches for each keypoint in the first image, and compute the detection rate as

$$\text{Detection Rate}(n) = 100 \cdot N_c(n)/N, \quad (10)$$

where  $N_c(n)$  is the number of keypoints from the first image, that have the correct match among the top  $n$  candidates.

Fig. 8-Top shows the detection rate for the synthetic dataset. Note that when no light changes are considered, the  $\text{GIH}[0.95, 1.05]$  yields almost identical results as our approach. However, its performance rapidly drops when photometric changes are introduced. In fact, even the  $\text{GIH}[0.25, 1.75]$  does not provide satisfactory results, showing that increasing the representation capabilities of the GIH is not necessarily correlated with a higher discrimination power. DaLI also consistently outperforms SIFT and DAISY, which are not specifically designed to tackle non-



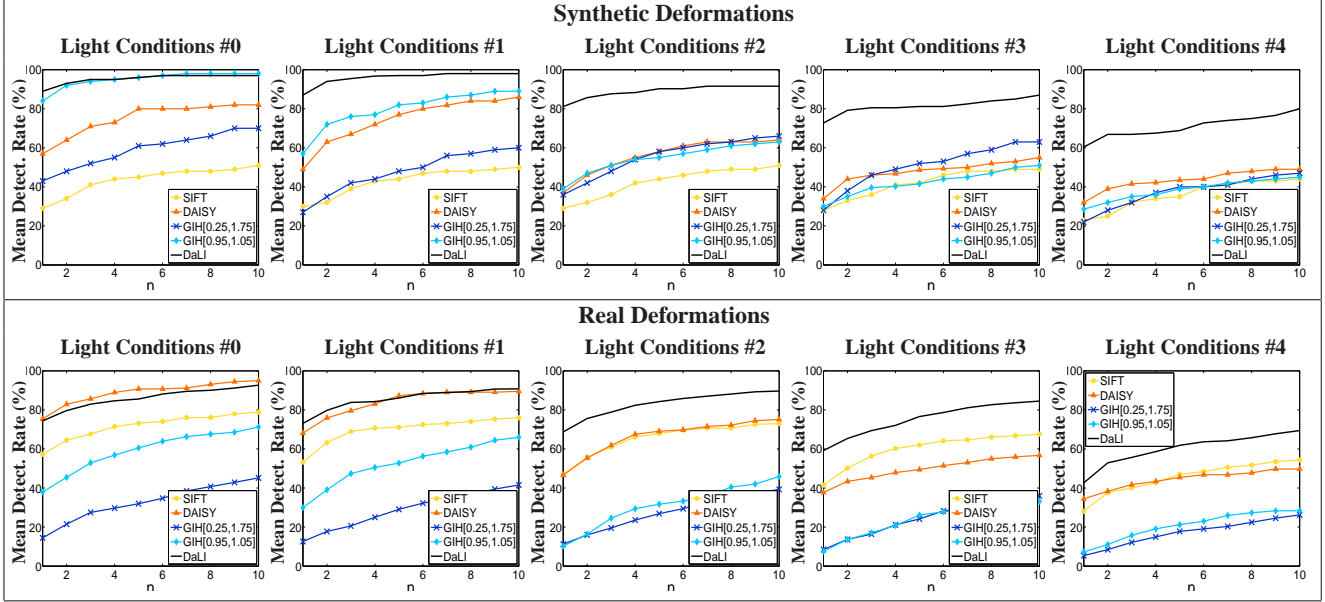


Figure 8. Mean detection rate on the synthetic and real deformation data sets, and for increasing amounts of light change.

linear deformations as those we evaluate in the synthetic experiments.

Under real deformations (Fig. 8-Bottom), the DAISY and GIH approaches change their roles. From one hand, when no light change is considered, DAISY performs almost equally as our approach. This is because most of the deformations of our real dataset can be locally approximated by affine transformations, for which both DAISY and SIFT have been proven to be robust. These methods can also tolerate certain photometric changes, but much less than DaLI. On the other hand, the GIH performs quite poorly for real deformations, even when no light changes are considered. This reveals another limitation of this approach, in that it assumes the effect of deformations is to locally change the position of image pixels, while in real deformations some of the pixels may disappear due to occlusions. Although our approach does not explicitly address occlusions, we can partially handle them by weighing the contribution of the pixels within each patch, by a function of the distance to the center. This makes that most of the information of our descriptor is concentrated in a small region surrounding the point of interest, and hence, being less sensitive to occlusions. Fig. 9 shows the results for a few pairs of input images. Observe how DaLI clearly outperforms the other approaches.

An finally, in Fig. 10 we compare the methods when only the best candidate is taken, that is, considering  $n = 1$  in Eq. 10. In addition, since our dataset contains repetitions of the same experiments with increasing amounts of deformation, we split the analysis into small and large deformations. Again, we show that our approach clearly does better than other methods, especially under illumination changes.

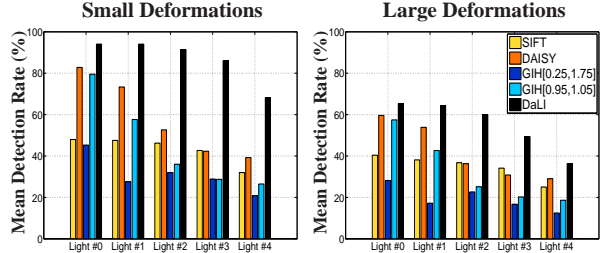


Figure 10. Detection Rates for small and large deformations, considering just the best candidate ( $n = 1$ ).

## 5. Conclusion and Future Work

Heat diffusion theory has been recently shown effective for 3D shape recognition tasks. In this paper, we have proposed using these tools to build the DaLI, a feature point descriptor for 2D image patches, that is invariant to non-rigid deformations and light changes. The advantages of our method with respect to state-of-the-art have been demonstrated by extensively testing on images containing both real and synthetic deformations.

There are still some issues that may be improved and are left for future work. For instance, we are currently representing the descriptor as a 3D array defined on the spatial and frequency domains. As a future work we will investigate different techniques for compacting this representation to allow for faster matching while retaining recognition rates. In addition, the technique we use to handle in-plane rotations and scaling searches within a wide range of values. This step can be readily speeded up, by incorporating prior orientation and scale estimation into our descriptor, as it is done for the SIFT descriptor.

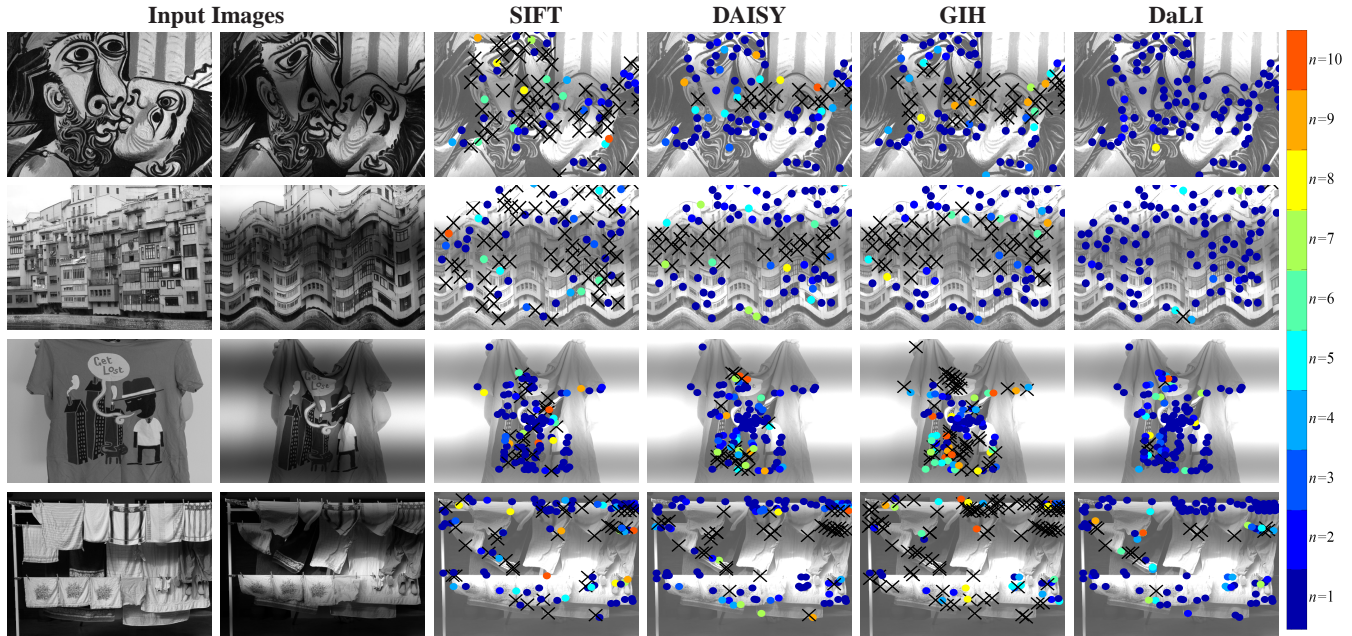


Figure 9. Matching results for a few samples from the test set. As in Fig. 1 the color of the circles indicates the position  $n$  of the correct match among the top candidates. If  $n > 10$  we consider the point as unmatched and mark it with a cross.

## References

- [1] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *ECCV*, 2006.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 2002.
- [3] A. Berg, T. Berg, J. Malik. Shape matching and object recognition using low distortion correspondences. In *CVPR*, 2005.
- [4] A. Bronstein, M. Bronstein, A. Bruckstein, and R. Kimmel. Analysis of two-dimensional non-rigid shapes. *IJCV*, 2007.
- [5] M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*, 2010.
- [6] M. Calonder, V. Lepetit, and P. Fua. BRIEF: Binary robust independent elementary features. In *ECCV*, 2010.
- [7] I. Chavel. *Eigenvalues in Riemannian Geometry*. London Academic Press, 1984.
- [8] H. Cheng, Z. Liu, N. Zheng, and J. Yang. A deformable local image descriptor. In *CVPR*, 2008.
- [9] M. Cho, J. Lee, and K. Lee. Feature correspondence and deformable object matching via agglomerative correspondence clustering. In *ICCV*, 2009.
- [10] Y. Ke and R. Sukthankar. PCA-SIFT: a more distinctive representation for local image descriptors. In *CVPR*, 2004.
- [11] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV*, 2005.
- [12] B. Levy. Laplace-beltrami eigenfunctions towards an algorithm that “understands” geometry. In *ICSMA*, 2006.
- [13] H. Ling and D. Jacobs. Deformation invariant image matching. In *ICCV*, 2005.
- [14] H. Ling and D. Jacobs. Shape classification using the inner-distance. *PAMI*, 2007.
- [15] H. Ling, X. Yang, and L. Latecki. Balancing deformability and discriminability for shape matching. In *ECCV*, 2010.
- [16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.
- [17] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 10(27):1615–1630, 2005.
- [18] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics*, 1993.
- [19] M. Reuter, F. Wolter, and N. Peinecke. Laplace-beltrami spectra as ‘shape-dna’ of surfaces and solids. *CAD*, 2006.
- [20] R. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *ESGP*, 2007.
- [21] J. Sanchez, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In *CVPR*, 2010.
- [22] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *Trans. on Image Processing*, 1998.
- [23] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Eurographics Symp. on Geometry Processing*, 2009.
- [24] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *PAMI*, 2010.
- [25] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *ECCV*, 2008.
- [26] A. Vedaldi and S. Soatto. Features for recognition: Viewpoint invariance for non-planar scenes. In *ICCV*, 2005.
- [27] G. Xu. Discrete laplace-beltrami operators and their convergence. *Computer Aided Geometric Design*, 2004.
- [28] S. Yanowitz and A. Bruckstein. A new method for image segmentation. *Comp. Vision, Graph. and Image Proc.*, 1989.
- [29] A. Yezzi. Modified curvature motion for image smoothing and enhancement. *Trans. on Image Processing*, 1998.